

A Five-Dimensional Framework for Authentic Assessment

Citation for published version (APA):

Gulikers, J. T. M., Bastiaens, T. J., & Kirschner, P. A. (2004). A Five-Dimensional Framework for Authentic Assessment. *Etr&D-Educational Technology Research and Development*, 52(3), 67-86.
<https://doi.org/10.1007/BF02504676>

DOI:

[10.1007/BF02504676](https://doi.org/10.1007/BF02504676)

Document status and date:

Published: 01/01/2004

Document Version:

Publisher's PDF, also known as Version of record

Document license:

CC BY

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

<https://www.ou.nl/taverne-agreement>

Take down policy

If you believe that this document breaches copyright please contact us at:

pure-support@ou.nl

providing details and we will investigate your claim.

Downloaded from <https://research.ou.nl/> on date: 05 May. 2023

Open Universiteit
www.ou.nl





Un marco de referencia de cinco dimensiones para la evaluación auténtica

Judith T. M. Gulikers
Theo J. Bastiaens
Paul A. Kirschner

Título original: A Five-Dimensional Framework for Authentic Assessment*. *ETR&D*, Vol. 52, No. 3, 2004, pp. 67-86 ISSN 1042-1629

Permiso de traducción de las páginas 67 – 76 de la Association for Educational Communications & Technology a través de Copyright Clearance Center, número de confirmación 2960199, septiembre de 2010, para Aseguramiento de la Calidad en la Educación y en el Trabajo.

La autenticidad es un elemento importante para los nuevos modelos de evaluación. El problema es que no es claro lo que significa la evaluación auténtica. En este artículo revisamos en primer lugar, la literatura sobre la autenticidad de las evaluaciones, junto con un marco de referencia de cinco dimensiones para diseñar evaluaciones auténticas con la práctica profesional como punto de partida. Después presentamos los resultados de un estudio cualitativo para determinar si el marco de referencia y si la importancia relativa de las cinco dimensiones está presente en las percepciones de los estudiantes y profesores de una escuela técnica de enfermería. Se discuten las implicaciones del marco de referencia, junto con temas importantes que deben considerarse al diseñar evaluaciones auténticas.

* Nota del traductor: el término assessment se tradujo al español por evaluación. En inglés tiene el sentido del proceso de valoración y se distingue del término evaluation que se aplica para la asignación del valor, sin embargo en español se utiliza el término evaluación para los dos casos.

En general se reconoce que para cumplir con las metas de la educación es necesario contar con una alineación constructiva entre la instrucción, el aprendizaje y la evaluación (ILA, por sus siglas en inglés) (Biggs, 1996). Un ejemplo de dicha alineación es la instrucción al frente de salón de clases en la que se aprenden hechos y que se evalúa mediante respuestas cortas o instrumentos de opción múltiple. Las prácticas ILA dentro de este tipo de educación pueden caracterizarse de la siguiente manera: enfoque instruccional: transmisión de conocimientos; enfoque de aprendizaje: memorización; y procedimiento de evaluación: exámenes estandarizados (Birenbaum, 2003). Este enfoque sobre la valoración también se conoce como la cultura de los exámenes (Birenbaum y Dochy, 1996) y consiste principalmente en instrumentos descontextualizados diseñados con principios psicométricos en formato de opción múltiple para evaluar la adquisición de conocimientos y de habilidades cognitivas de bajo nivel. Los exámenes se usan principalmente de manera sumativa para diferenciar entre estudiantes y clasificarlos de acuerdo con su logro. Sin embargo, la alineación compatible con las metas educativas actuales ha cambiado a lo largo de los años. Las metas educativas actuales se enfocan en el desarrollo de estudiantes y futuros empleadores competentes más que en la mera adquisición de conocimientos. Las prácticas ILA que caracterizan a estas metas son: enfoque instruccional: énfasis en el desarrollo de aprendizajes y competencias; enfoque de aprendizaje: construcción de conocimiento reflexiva y activa; y procedimiento de evaluación: evaluación contextualizada e interpretativa del desempeño. Aquí, el objetivo de la evaluación es la adquisición de competencias y de procesos de pensamiento de alto nivel en lugar de conocimientos fácticos y de habilidades básicas. La función de dicha evaluación cambia de ser sumativa a ayudar a lograr el objetivo formativo de promover y mejorar el aprendizaje de los estudiantes. Esta visión requiere evaluaciones alternativas debido a que los exámenes estandarizados de opción múltiple no son adecuados para el enfoque (Birenbaum y Dochy, 1996; Segers, Dochy y Cascallar, 2003). Birenbaum y Dochy (1996) caracterizaron las evaluaciones alternativas de la siguiente manera: los estudiantes tienen la responsabilidad de su propio aprendizaje, reflexionan, colaboran y establecen un diálogo continuo con sus profesores. La evaluación incluye tareas y contextos auténticos o de la vida real, así como diferentes momentos y métodos de evaluación que permitan establecer una calificación para determinar el aprendizaje o el desarrollo de los estudiantes. Se espera que aumentar la autenticidad de una evaluación tenga una influencia positiva en el aprendizaje y la motivación de los estudiantes (Herrington y Herrington, 1998). Sin embargo, la autenticidad es sólo una dimensión de la evaluación descrita de manera somera ya que se le considera un concepto familiar y generalmente común que no necesita definirse de manera explícita (Petraglia, 1998). Este artículo se centra en definir la autenticidad en la evaluación basada en competencias sin ignorar la importancia de otras características de la evaluación alternativa.

Con base en un extenso estudio de la literatura, se presenta un marco de referencia teórico que consiste en cinco dimensiones de la evaluación que pueden variar en el grado de autenticidad. Después de describir este marco de referencia, se explican los resultados de un estudio cualitativo. Este estudio exploró si el marco de referencia constituye una descripción completa de la autenticidad o si carece de elementos importantes, así como la importancia relativa de las dimensiones en la percepción de los estudiantes y profesores en una escuela de enfermería.

La importancia de la evaluación auténtica basada en competencias

Las dos razones más importantes para utilizar evaluaciones auténticas basadas en competencias son (a) su validez de constructo y (b) su impacto en el aprendizaje de los estudiantes, llamado también validez consecuente (Gielen, Dochy y Dierrick, 2003). La *validez de constructo* de una evaluación se relaciona con el grado en que mide lo que debe medir. Con respecto a la evaluación de competencias, esto significa que (a) las tareas deben de reflejar de manera apropiada la competencia a evaluar, (b) el contenido de la evaluación incluye tareas auténticas que representan problemas de la vida real del campo de conocimiento que se evalúa y (c) los procesos de pensamiento que los expertos utilizan para resolver problemas reales también constituyen un requisito de la tarea de evaluación (Gielen et al., 2003). Con base en estos criterios, las evaluaciones auténticas basadas en competencias tienen una mayor validez de constructo que los llamados exámenes objetivos o tradicionales para medir las competencias.

La validez consecuente describe tanto los efectos pretendidos como los no pretendidos que la evaluación tiene sobre la instrucción o la enseñanza (Bibbs, 1996) y sobre el aprendizaje estudiantil (Dochy y McDowell, 1998).

Como se ha manifestado, la teoría de Biggs (1996) sobre la alineación constructiva enfatiza que la educación efectiva requiere que la instrucción, el aprendizaje y la evaluación sean compatibles. Si los estudiantes perciben un desfase entre los mensajes de la instrucción y la evaluación, es difícil lograr un impacto positivo en el aprendizaje de los estudiantes (Segers, Dierick y Dochy, 2001). Investigadores como Frederiksen (1984, "The Real Test Bias"), Prodromou (1995, "Backwash Effect"), Gibbs (1992, "Tail Wags the Dog") y Sambell y McDowell (1998, "Hidden Curriculum") han corroborado el impacto de la evaluación sobre la instrucción y el aprendizaje de los estudiantes. Frederiksen y Prodromou implican que los exámenes tienen una fuerte influencia sobre lo que se enseña debido a que los profesores enseñan para el examen aunque éste se enfoque en temas que los profesores no consideran tan importantes. Gibbs enfatizó que el aprendizaje estudiantil depende en gran medida de la evaluación y de las percepciones que el estudiante tiene sobre los requisitos de evaluación. Sambell y McDowell sostienen que los efectos que la instrucción y la evaluación tienen sobre el aprendizaje se basan en gran medida en las percepciones de los estudiantes y profesores sobre el currículum, y pueden desviarse de sus intenciones reales. Las cuatro ideas apoyan la propuesta de que el aprendizaje y la evaluación son dos caras de la misma moneda y que tienen una fuerte influencia una sobre la otra. Para hacer que el aprendizaje estudiantil camine en la dirección del desarrollo de competencias, se necesita una instrucción auténtica basada en competencias alineada con una evaluación auténtica basada en competencias.

Definición de la evaluación auténtica

La pregunta es: ¿Qué es la autenticidad? Los investigadores tienen diversas opiniones sobre la autenticidad. Algunos consideran que la evaluación auténtica es sinónimo de evaluación del desempeño (Hart, 1994; Torrance, 1995), mientras que otros argumentan que la evaluación auténtica tiene un énfasis especial en el valor real de la tarea y del contexto (Herrington y Herrington, 1998). Reeves y Okey (1996) señalan que la diferencia crucial entre evaluación del desempeño y la evaluación auténtica es el grado de fidelidad de la tarea y de las condiciones bajo las cuales debería ocurrir de manera normal el desempeño. La evaluación auténtica se enfoca que una alta *fidelidad*, mientras que esto no es un factor tan importante en la evaluación del desempeño. *Estas distinciones entre evaluación auténtica y de desempeño* indican que cada evaluación auténtica es una evaluación de desempeño, pero no al contrario (Meyer, 1992).

Savery y Duffy (1995) definen la autenticidad de una evaluación como la *similitud* entre las demandas cognitivas (el pensamiento requerido) de la evaluación y las demandas cognitivas en la situación en la que se basa la evaluación. Una situación definida mediante criterios refleja o simula una situación de la vida real con la que los estudiantes podrían encontrarse durante una pasantía o en su vida profesional futura. Darling-Hammond y Snyder (2000) argumentan que considerar sólo el pensamiento requerido es demasiado estrecho. En su opinión, los estudiantes necesitan desarrollar competencias debido a que la vida real requiere que tengan la capacidad de integrar y coordinar conocimientos, habilidades y actitudes y la capacidad para aplicarlas a nuevas situaciones (van Merriënboer, 1997). Birenbaum (1996) aclara el concepto de competencia haciendo énfasis en que los estudiantes necesitan desarrollar no sólo competencias cognitivas como la resolución de problemas y el pensamiento crítico, sino también competencias metacognitivas como la reflexión y competencias sociales como la comunicación y la colaboración.

La definición de *evaluación auténtica* utilizada en este estudio es: una evaluación que requiere que los estudiantes apliquen las mismas competencias o combinación de *conocimientos, habilidades y actitudes* tanto en la situación definida mediante criterios como en la vida profesional. El nivel de autenticidad de una evaluación se define, por lo tanto por su grado de semejanza a la situación ficticia. Esta idea se extiende y se especifica en el marco teórico de referencia que describe que una evaluación puede asemejarse a una situación definida mediante criterios en varias dimensiones.

Un factor que complica la cuestión es el hecho de que la autenticidad es subjetiva (Honebein, Duffy y Fishman, 1993; Huang, 2002; Petraglia, 1998) y que depende de la percepción. Esto implica que lo que los estudiantes, los profesores y los desarrolladores de evaluaciones perciben como auténtico no es necesariamente lo mismo. Si estas percepciones difieren, entonces el hecho de que los profesores normalmente desarrollan evaluaciones auténticas de acuerdo con su propia visión ocasiona un problema: aunque hagamos nuestro mayor esfuerzo para desarrollar evaluaciones auténticas, puede ser inútil si el estudiante no las percibe como tales. Puede interpretarse que en este proceso, conocido como *preautenticación* (Huang, 2002; Petraglia, 1998) sea imposible diseñar una evaluación auténtica o que sea muy importante examinar de manera cuidadosa las experiencias de otros usuarios sobre las evaluaciones auténticas antes de diseñar dicho tipo de evaluaciones (Nicaise, Gibney y Crane, 2000). Nosotros nos inclinamos por la segunda interpretación.

Esta discusión sobre la evaluación auténtica y la validez muestra que:

1. A la luz de la teoría de alineación constructiva (Biggs, 1996) la evaluación auténtica debería estar alineada con la instrucción para poder influir de manera positiva en el aprendizaje de los estudiantes.
2. La evaluación auténtica requiere que los estudiantes demuestren competencias relevantes mediante un logro significativo y que valga la pena (Resnick, 1987; Wiggins, 1993).
3. La autenticidad es subjetiva, lo que hace que las percepciones de los estudiantes sean importantes para que la evaluación auténtica logre influir en el aprendizaje.

Estos tres elementos generaron el siguiente marco general de referencia (Figura 1) para la evaluación auténtica en las prácticas educativas.

El concepto de *logro auténtico*, como lo utilizamos aquí, requiere una explicación. Este artículo trata sobre la evaluación auténtica en general, sin importar el nivel o campo de estudios. Esto no significa que desechemos el concepto de logro académico auténtico (Newmann, 1997), sino que lo consideramos como un subconjunto específico dentro de un campo de estudios específico, a saber, convertirse en un académico. En esto, coincidimos con Brown, Collins y Duguid (1998) quienes de la misma manera consideran al logro auténtico más amplio que el logro *académico* auténtico.

En la siguiente sección se explican las cinco dimensiones (el marco teórico de referencia) que pueden variar en su grado de autenticidad al determinar la autenticidad de una evaluación. El propósito de este marco de referencia es esclarecer el concepto de autenticidad de las evaluaciones y ofrecer lineamientos para la implementación de elementos de autenticidad en las evaluaciones basadas en competencias.

Hacia un marco de referencia de cinco dimensiones para la evaluación auténtica.

Para definir la evaluación auténtica, realizamos una revisión de la literatura sobre la evaluación auténtica, sobre la autenticidad y la evaluación en general y sobre las percepciones de los estudiantes sobre algunos elementos de evaluación (auténtica). Se identifican cinco dimensiones de la evaluación auténtica: (a) la tarea de evaluación, (b) el contexto físico, (c) el contexto social, (d) el resultado o forma de evaluación y (e) los criterios de evaluación. Estas dimensiones pueden variar en su nivel de autenticidad (son continuas). Es una concepción errónea pensar que algo es o auténtico o no auténtico (Cronin, 1993; Newmann y Wehlage, 1993) debido a que el nivel de autenticidad no es solamente una característica de la evaluación seleccionada; es necesario definirla en relación con la situación definida mediante criterios derivada de la práctica profesional. Por ejemplo: realizar una evaluación en equipo sólo es auténtico si la tarea de evaluación seleccionada también se realiza en equipo en la vida real. El punto principal del marco de referencia es que cada una

de las cinco dimensiones puede asemejarse a la situación definida mediante criterios en diferente grado, aumentando o disminuyendo así la autenticidad de la evaluación.

Debido a que la evaluación definida mediante criterios debería alinearse a la instrucción auténtica (Biggs, 1996, van Merriënboer, 1997), las cinco dimensiones de un marco de referencia para la evaluación auténtica también se aplican a la instrucción auténtica. Aunque el enfoque de este artículo es la evaluación auténtica, se incluye una interpretación de las cinco dimensiones para la instrucción auténtica con el fin de mostrar que las mismas cinco dimensiones pueden utilizarse para lograr una alineación entre la instrucción auténtica y la evaluación auténtica. Las dimensiones y los elementos subyacentes de la instrucción auténtica que se presentan en la Figura 2 y en la Figura 3 funcionan también para la evaluación auténtica.

Como se muestra en las figuras, las tareas de aprendizaje y las tareas de evaluación son muy semejantes. Esto es lógico, pues la tarea de aprendizaje estimula a los estudiantes para que desarrollen las competencias que tienen los profesionales y las tareas de evaluación requieren que los estudiantes demuestren tener esas mismas competencias sin un apoyo adicional (van Merriënboer, 1997). Schnitzer (1993) enfatiza que para que la evaluación auténtica sea efectiva, es necesario que los estudiantes tengan la oportunidad de practicar con la forma de evaluación antes de que se utilice como evaluación. Esto implica que la tarea de aprendizaje debe parecerse a la tarea de evaluación y sólo difiera en las metas subyacentes. Las tareas de aprendizaje son para aprender y las tareas de evaluación para evaluar el nivel de aprendizaje del estudiante para mejorar (formativo) o para tomar decisiones (sumativo). Estos modelos muestran que un marco de referencia de cinco dimensiones puede manejar una alineación (conceptual) entre la instrucción auténtica y la evaluación auténtica. La interpretación y la evaluación de las cinco dimensiones para la evaluación auténtica se explica y se revisa con más profundidad en el resto de este artículo.

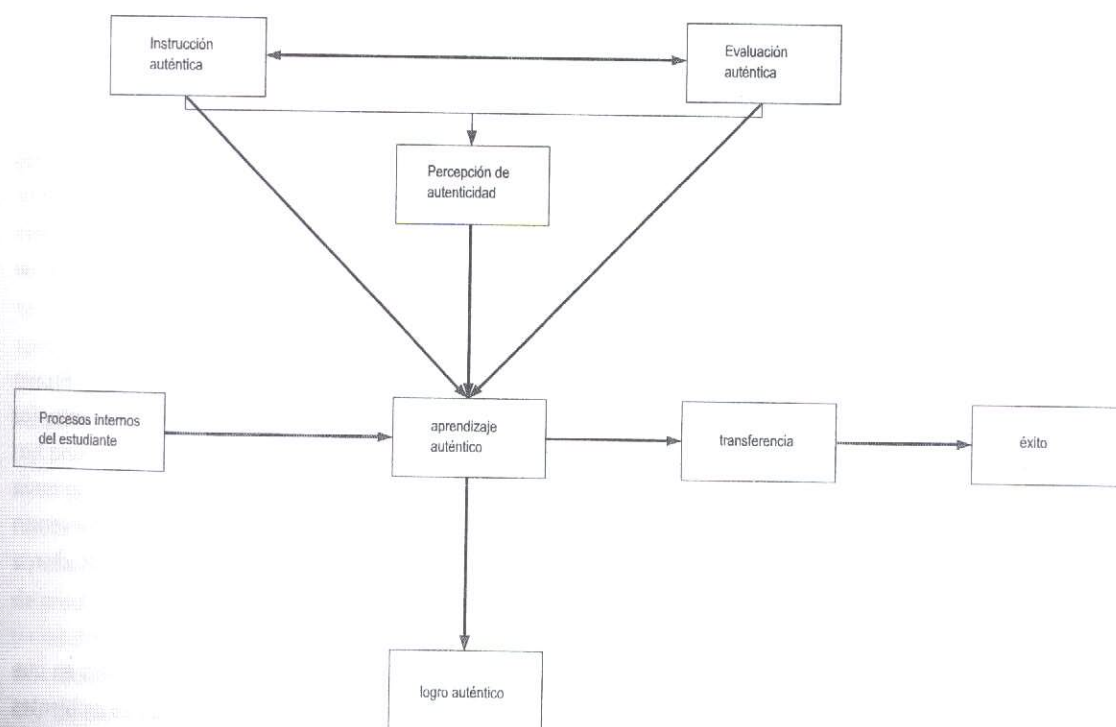


Figura 1 Marco general de referencia

Una argumentación a favor de las cinco dimensiones de la evaluación auténtica.

Como se ha mencionado, existe una confusión y existe una gran diferencia de opiniones sobre lo que la autenticidad de las evaluaciones es en realidad y sobre qué elementos de la evaluación son importantes para la autenticidad. Para tratar de esclarecer esta situación, se revisó la literatura y así se explican las diferentes ideas sobre la autenticidad. Surgieron muchos subconceptos y sinónimos que se analizaron de manera conceptual y se dividieron en categorías obteniendo como resultado los cinco aspectos principales de la autenticidad. La noción de autenticidad como continuo (Mewmann y Wehlage, 1993) dio por resultado la conceptualización de estos cinco aspectos como dimensiones que pueden variar en su grado de autenticidad.

Tarea. Una tarea auténtica es una tarea problema que confronta al estudiante con actividades que también se realizan en la práctica profesional. El hecho de que una tarea auténtica sea crucial para la evaluación auténtica es indiscutible (Herrington y Herrington, 1998; Newmann, 1997; Wiggins, 1993), pero los diversos investigadores hacen énfasis en diferentes elementos de una *tarea auténtica*. Nuestro marco de referencia define una tarea auténtica como una tarea que se asemeja a la tarea definida mediante criterios con respecto a la integración de conocimientos, habilidades y actitudes, a su complejidad y a su propiedad (ver Kirschner, Martens y Strijbos, 2004). Más aun, los usuarios de las tareas de evaluación deben percibir la tarea y los elementos descritos arriba como representativas, relevantes y significativas.

Una evaluación auténtica requiere que los estudiantes integren conocimientos, habilidades y actitudes de la misma manera que lo hacen los profesionales (van Merriënboer, 1997). Además, la tarea de evaluación, debe parecerse en *complejidad* a la tarea definida mediante criterios (Petraglia, 1998; Uhlenbeck, 2002). Esto no significa que cada tarea de evaluación deba ser muy compleja. Aunque la mayoría de problemas auténticos son complejos, implican la multidisciplinariedad, presentan una estructura poco definida y tiene varias soluciones posibles (Herrington y Herrington, 1998; Kirschner, 2002; Wiggins, 1993), los problemas de la vida real también pueden ser sencillos, bien estructurados, con una respuesta correcta y que requieran sólo una disciplina (Cronin, 1993). La misma necesidad de semejanza existe para la *apropiación* de la tarea y de los procesos de desarrollo de una solución. La apropiación de los estudiantes en la tarea de evaluación debe asemejarse a la apropiación de los profesionales en las tareas definidas mediante criterios. Savery y Duffy (1995) argumentan que otorgarles a los estudiantes la propiedad de la tarea y del proceso de desarrollar una solución es crucial para comprometerlos en el aprendizaje auténtico y en la resolución de problemas. Por otro lado, en la vida real, los empleadores imponen las actividades y los profesionales utilizan herramientas y procedimientos estandarizados para resolver un problema, disminuyendo así la propiedad del empleador. Por lo tanto, el marco teórico de referencia supone que para lograr que los estudiantes sean competentes al enfrentarse con problemas profesionales, la tarea de evaluación debe asemejarse en complejidad y propiedad a las situaciones de la vida real.

Hasta este punto, la autenticidad de la tarea parece ser una dimensión bastante objetiva. Sambell, McDowell y Brown (1997) comparten esta objetividad al mostrar que es crucial que los estudiantes perciban la tarea como relevante, (a) que perciban el vínculo con una situación de la vida real o de un puesto de trabajo; o (b) la consideren como una habilidad transferible valiosa. McDowell (1995) también enfatiza que los estudiantes deben ver un vínculo entre la tarea de evaluación y sus intereses personales antes de que perciban la tarea como significativa. Es claro que la relevancia o significatividad relativa diferirá de un estudiante a otro e incluso posiblemente cambie a medida que el estudiante adquiere experiencia.

Contexto físico. El lugar en el que nos encontramos casi siempre, si no es que siempre, determina la manera en que hacemos las cosas y casi siempre también el lugar real es más sucio (literal y figurativamente) que los seguros ambientes de aprendizaje. Piense, por ejemplo en una evaluación de mecánica de automóviles para los militares. La capacidad de un soldado para encontrar el problema en un vehículo que esté fuera de servicio puede evaluarse en un garaje limpio con todo el equipo imaginable a su disposición, pero un ambiente físico futuro puede implicar una zona de guerra, con-

diciones climáticas adversas, menos espacio y menos equipo. Aunque la tarea en sí sea auténtica, puede cuestionarse si evaluar a los estudiantes en un ambiente limpio y seguro en verdad determina su habilidad para usar sus competencias en situaciones reales.

El contexto físico de una evaluación auténtica debe reflejar la manera en que el conocimiento, habilidades y actitudes se utilizarán en la práctica profesional (Brown et al. 1989; Herrington y Oliver, 2000). En ocasiones la *fidelidad* se utiliza en el contexto de las simulaciones por computadora, lo que describe qué tan estrechamente la simulación imita a la realidad (Alessi, 1988). Con frecuencia, la evaluación auténtica se encuentra en contextos de alta fidelidad. La presentación de materiales y la cantidad de detalles presentados en el contexto son aspectos importantes del grado de fidelidad. De la misma manera, un importante elemento de la autenticidad del contexto físico es que el número y tipo de recursos disponibles (Segers, Dochy y De Corte, 1999) representan información tanto relevante como irrelevante. Por ejemplo, Resnick (1987) argumenta que la mayoría de los exámenes escolares involucran trabajo de memoria, mientras que las actividades extramuros con frecuencia están íntimamente ligadas con herramientas y recursos (calculadoras, tablas, estándares) que ocasionan que los exámenes escolares sean menos auténticos. Segers et al. (1999) argumentan que no sería auténtico privar a los estudiantes de los recursos, ya que los profesionales sí tienen acceso a ellos. Otra característica importante que es crucial para lograr un contexto físico auténtico es el tiempo que los estudiantes dedican a desempeñar la tarea de evaluación (Wiggins, 1989). Normalmente, los exámenes se aplican en un lapso restringido, por ejemplo, dos horas dedicadas por completo al examen. En la vida real, las actividades profesionales implican más tiempo distribuido a lo largo de diferentes días o, por el contrario, requieren una reacción rápida e inmediata en fracciones de segundo. Wiggins (1989) sostiene que una evaluación auténtica no debe depender de limitaciones temporales arbitrarias y poco realistas. En resumen, el nivel de autenticidad del contexto físico se define por la semejanza de los elementos con la situación definida mediante criterios.

Contexto social. No sólo el contexto físico, sino que también el contexto social influye en la autenticidad de la evaluación. En la vida real, el trabajo en equipo es muy común y Resnick (1987) enfatiza que el aprendizaje y el desempeño fuera de la escuela se realizan en su mayor parte dentro de un sistema social. Por lo tanto, un modelo para la evaluación auténtica debería considerar los procesos sociales presentes en los contextos de la vida real. Lo que es muy importante en una evaluación auténtica es que los procesos sociales de la evaluación se asemejen a los procesos sociales en una situación equivalente en la realidad. En este punto, este marco de referencia no está de acuerdo con la literatura sobre la evaluación auténtica que define a la colaboración como una característica de la autenticidad (Herrington y Herrington, 1998). En nuestro marco de referencia se argumenta que si la situación real requiere de una colaboración la evaluación debe también incluir una colaboración, pero sí normalmente la situación se maneja de manera individual, la evaluación también debe hacerse de manera individual. Cuando la evaluación requiere colaboración, algunos procesos como la interacción social, la interdependencia positiva y la rendición individual de cuentas deben tenerse en cuenta (Salvin, 1989). Sin embargo, cuando la evaluación es individual, el contexto social debe favorecer algún tipo de competencia entre los estudiantes.

Resultado o forma de la evaluación. Una evaluación incluye una asignación de la evaluación (dentro de contextos físicos y sociales determinados) que genera un resultado de evaluación, que a su vez es evaluado contra ciertos criterios de valoración (Moerkerke, Doorten y de Roode, 1999). El resultado de la evaluación se relaciona con el tipo y la cantidad de productos de la tarea de evaluación independientemente del contenido de la evaluación. En el marco de referencia, un resultado o forma auténticos se caracterizan por contar con cuatro elementos. Debe ser un (a) producto o desempeño de calidad que puede pedirse a los estudiantes en la vida real (Wiggins, 1989). Este producto o desempeño debe ser una (b) demostración que permita hacer inferencias válidas sobre las competencias subyacentes (Darling-Hammond y Snyder, 2000). Debido a que la demostración de competencias relevantes no siempre es posible en un solo examen, una evaluación auténtica debe incluir (c) una gama de tareas y múltiples indicadores del aprendizaje para poder llegar a conclusiones válidas (Darling-Hammond y Snyder, 2000). Uhlenbeck (2002) demuestra que una combinación de diferentes

métodos de evaluación cubre de manera adecuada el rango completo de comportamientos de la enseñanza profesional. Por último, los estudiantes deben (d) presentar su trabajo a otras personas, ya sea de forma oral o escrita debido a que es importante que defiendan su trabajo para asegurar que su dominio aparente sea genuino (Wiggins, 1989).

Criterios y estándares. Los *criterios* son aquellas características del resultado de la evaluación que se evalúan; los *estándares* son el nivel de desempeño esperado de estudiantes de diferentes grados y edades (Arter y Spandel, 1992). En una evaluación auténtica es importante establecer criterios y hacerlos explícitos y transparentes para los estudiantes de manera inicial ya que esto guía el aprendizaje (Sluijsmans, 2002) y después de todo, en la vida real, los empleados saben bajo qué criterios se juzgará su desempeño. Esto implica que la evaluación auténtica requiere de un juicio que haga referencia a criterios. Además, algunos criterios deben estar relacionados con un resultado realista y deben explicar las características o requisitos del producto, desempeño o solución que los estudiantes deben crear. Además, los criterios y estándares deben estar relacionados con el desarrollo de competencias profesionales y deben estar basados en criterios utilizados en situaciones de la vida real (Darling, Hammond y Snyder, 2000).

Además de constituir la base de los criterios de la situación definida mediante criterios en la vida real, los criterios de una evaluación auténtica también pueden estar basados en la interpretación de las otras cuatro dimensiones del marco de referencia. Por ejemplo, si el contexto físico determina que la evaluación auténtica de una competencia requiere cinco horas, uno de los criterios debe de ser que los estudiantes necesitan producir un resultado dentro de un periodo de cinco horas. Por otro lado, los criterios que se basan en la práctica profesional también pueden guiar la interpretación de las otras cuatro dimensiones de la evaluación auténtica. En otras palabras, el marco de referencia aboga por una relación recíproca entre la dimensión de criterios y las otras cuatro dimensiones.

Algunas consideraciones

¿Qué significa todo esto en el momento en el que los profesores o diseñadores instruccionales intentan desarrollar evaluaciones auténticas? ¿Qué necesitan considerar?

La primera consideración tiene que ver con la validez predictiva. Si se persigue el objetivo educativo de desarrollar empleados competentes, aumentar la autenticidad de una evaluación será favorable. Es factible que un mayor grado de autenticidad aumente la validez predictiva de la evaluación debido a la semejanza entre la evaluación y la práctica profesional real. Sin embargo, no debemos eliminar los aspectos positivos junto con los negativos al intentar lograr una mejora. Los exámenes objetivos todavía son muy útiles para ciertos propósitos como evaluaciones sumativas de alto impacto de los logros individuales, en no es la intención predecir la habilidad futura del estudiante para funcionar de manera competente en la práctica profesional.

Otra consideración para el diseño de la evaluación auténtica es que no debemos de perder de vista el nivel educativo de los estudiantes. Puede ser que los estudiantes de los niveles más bajos no puedan dominar la autenticidad de una situación profesional real y compleja. Si se les obliga a hacerlo puede ocasionar una sobrecarga cognitiva y por lo tanto tener un impacto negativo en el aprendizaje (Sweller, van Merriënboer y Paas, 1998). Como resultado, una situación definida mediante criterios tendrá que ser una abstracción de la práctica profesional real para que los estudiantes de un cierto nivel educativo puedan realizarla. La pregunta que surge inmediatamente en este contexto es: ¿Cómo se crea una evaluación auténtica para estudiantes que no están preparados para funcionar como profesionales? La respuesta es que la autenticidad de una evaluación debe definirse por su grado de semejanza con la situación definida mediante criterios (es decir, una abstracción de la práctica profesional) y no necesariamente con la práctica profesional real. Van Merriënboer (1997) argumenta que una abstracción de la práctica profesional real (es decir, la situación definida mediante criterios) puede ser auténtica en la medida que la situación abstraída requiera que los estudiantes realicen la competencia com-

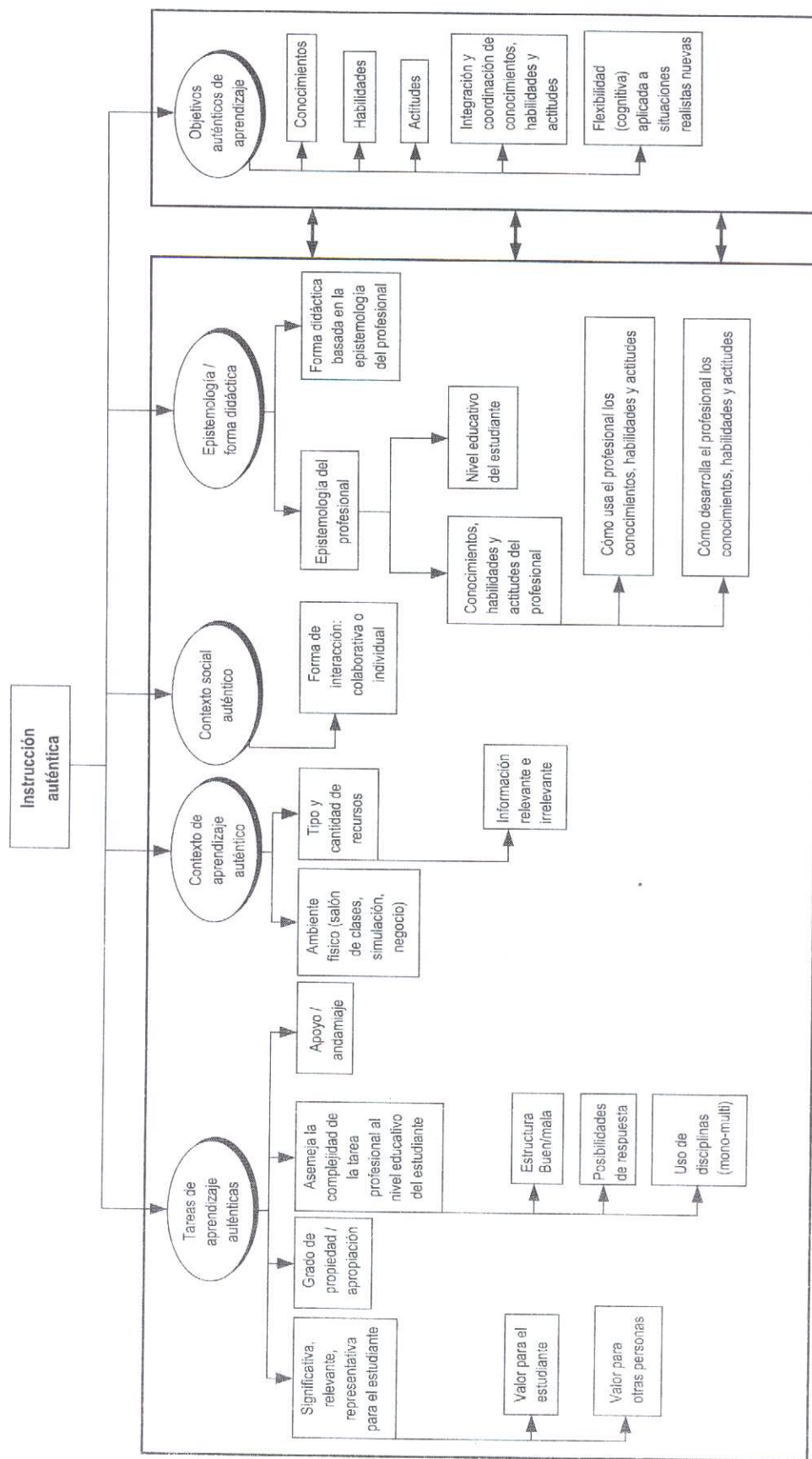


Figura 2 Modelo de cinco dimensiones para la instrucción auténtica

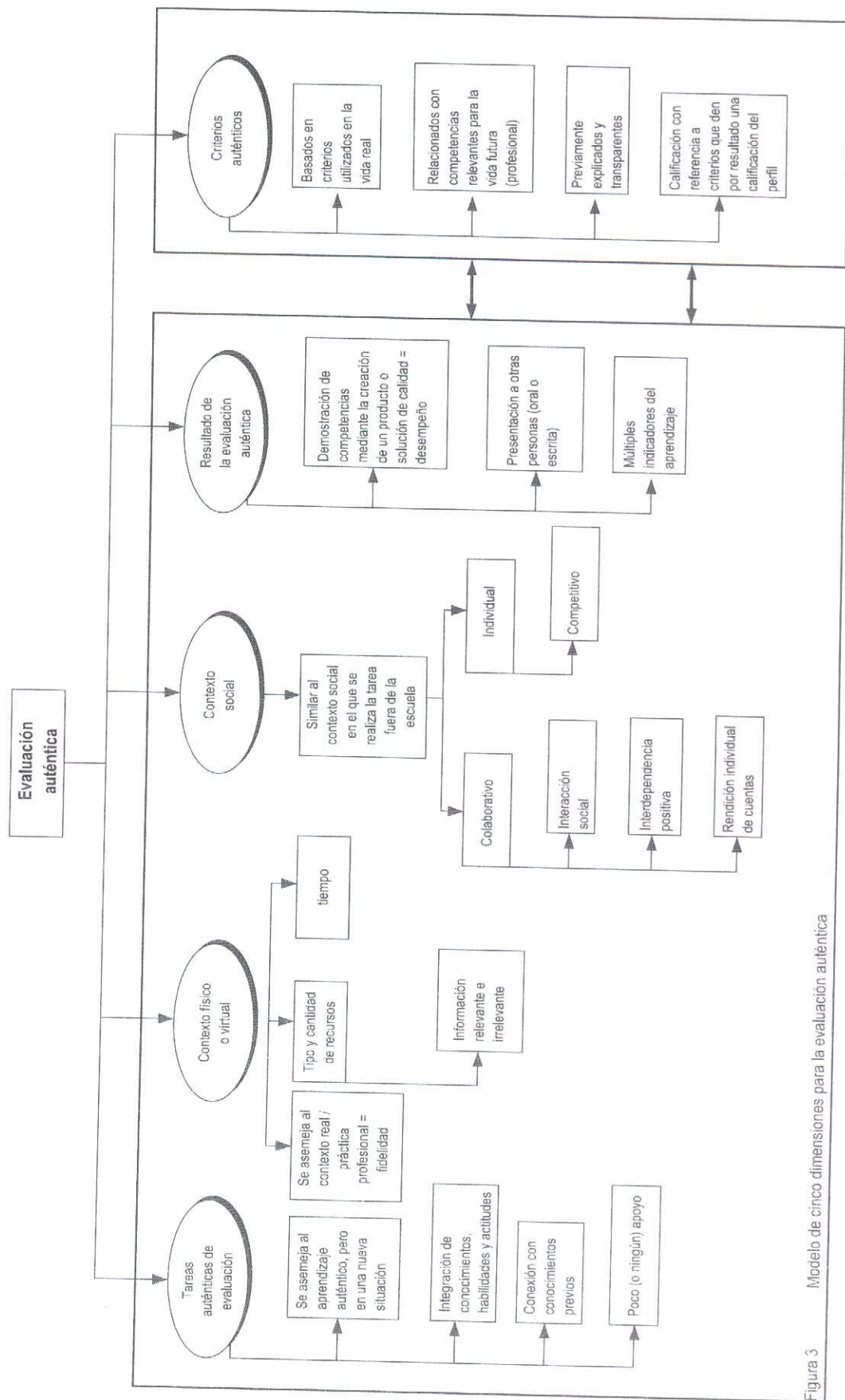


Figura 3 Modelo de cinco dimensiones para la evaluación auténtica

pleta como parte de un grupo de competencias constitutivas. La abstracción resulta de la simplificación de factores de contexto que complican el desempeño de la competencia completa.

Una tercera consideración también esclarece la cuestión planteada en las secciones previas, a saber, la subjetividad de la autenticidad. La percepción de lo que es la autenticidad puede cambiar como resultado del nivel educativo, de los intereses personales, de la edad o del grado de experiencia que se tenga en la práctica profesional (Honebein et al., 1993). Esto implica que las cinco dimensiones que se explican en el marco de referencia para la evaluación auténtica no son absolutas, sino variables. Es posible que evaluar la competencia profesional de los estudiantes en su último año de estudios, cuando probablemente hayan realizado pasantías y tengan una mejor idea de la práctica profesional, requiera un mayor grado de autenticidad del contexto físico que al evaluar a estudiantes de primer año, quienes en general tienen poca experiencia práctica. Los diseñadores deben considerar las perspectivas cambiantes de los estudiantes al diseñar una evaluación auténtica.

En el texto original continua la descripción del estudio que se llevó a cabo para validar las cinco dimensiones propuesta, p. 76 – 83

Judith T.M. Gulikers [judith.gulikers@ou.nl], Theo J. Bastiaens, y Paul A. Kirschner están en el Educational Technology Expertise Center en la Open University of the Netherlands, P.O. Box 2960, 6401 DL Heerlen, The Netherlands.

Los autores agradecen a Marijke Bijmens por su ayuda en organizar la participación de los profesores y alumnos en el GSS. Ellos también agradecen al Dr. Robert Schuur su apoyo para llevar a cabo las sesiones en el GSS.

BIBLIOGRAFÍA

Alessi, S. M. (1988). Fidelity in the design of instructional simulations. *Journal of Computer-Based Instruction*, 15(2), 40–47.

Arter, J. A., & Spandel, V. (1992). An NCME instructional module on: Using portfolio of student work in instruction and assessment. *Educational Measurement: Issues and Practice*, 11(1), 36–45.

Biggs, J. (1996). Enhancing teaching through constructive alignment. *Higher Education*, 32, 347–364.

Birenbaum, M. (1996). Assessment 2000: Towards a pluralistic approach to assessment. In M. Birenbaum & F. J. R. C. Dochy (Eds.), *Alternatives in assessment of achievements, learning processes and prior knowledge* (pp. 3–29). Boston, MA: Kluwer Academic Publishers.

Birenbaum, M. (2003). New insights into learning and teaching and their implications for assessment. In M. Segers, F. Dochy, & E. Cascallar (Eds.), *Optimising new modes of assessment: In search of quality and standards* (pp. 13–36). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Birenbaum, M., & Dochy, F. J. R. C. (1996). *Alternatives in assessment of achievements, learning processes and prior knowledge*. Boston, MA: Kluwer Academic Publishers.

Brown, J. S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher*, 18(1), 32–42.

-
- Cronin, J. F. (1993). Four misconceptions about authentic learning. *Educational Leadership*, 50(7), 78–80.
- Darling-Hammond, L., & Snyder, J. (2000). Authentic assessment in teaching in context. *Teaching and Teacher Education*, 16, 523–545.
- Dierick, S., Dochy, F., & Van de Watering, G. (2001). Assessment in het hoger onderwijs: Over de implicaties van nieuwe toetsvormen voor de edumetrie. [Assessment in higher education: About the implications of new assessment forms for edumetrics] *Tijdschrift voor Hoger Onderwijs*, 19(1), 2–18.
- Dochy, F. J. R. C., & McDowell, L. (1998). Assessment as a tool for learning. *Studies in Educational Evaluation*, 23(4), 279–298.
- Frederiksen, N. (1984). The real test bias, influences of testing and teaching on learning. *American Psychologist*, 39(3), 193–202.
- Gibbs, G. (1992). *Improving the quality of student learning*. Bristol, UK: Technical and Educational Services.
- Gielen, S., Dochy, F., & Dierick, S. (2003). The influence of assessment on learning. In M. Segers, F. Dochy, & E. Cascallar (Eds.), *Optimising new modes of assessment: In search of quality and standards* (pp. 37–54). Dordrecht, The Netherlands: Kluwer Academic Publishers.
- Hart, D. (1994). *Authentic assessment: A handbook for education*. Menlo Park, CA: Addison-Wesley Publishing Company.
- Herrington, J., & Herrington, A. (1998). Authentic assessment and multimedia: How university students respond to a model of authentic assessment. *Higher Educational Research & Development*, 17(3), 305–322.
- Herrington, J., & Oliver, R. (2000). An instructional design framework for authentic learning environments. *Educational Technology Research and Development*, 48(3), 23–48.
- Honebein, P. C., Duffy, T. M., & Fishman, B. J. (1993). Constructivism and the design of learning environments: Context and authentic activities for learning. In T. M. Duffy, J. Lowyck, & D. H. Jonassen (Eds.), *Designing environments for constructive learning* (pp. 88–108). Berlin: Springer-Verlag.
- Huang, H. M. (2002). Towards constructivism for adult learners in online learning environments. *British Journal of Educational Technology*, 33, 27–37.
- Kirschner, P. A. (2002). *Three worlds of CSCL: Can we support CSCL?* Heerlen: Open University of the Netherlands.
- Kirschner, P. A., Martens, R. L., & Strijbos, J. W. (2004). CSCL in higher education? A framework for designing multiple collaborative environments. In P. Dillenbourg (Series Ed.) & J. W. Strijbos, P. A. Kirschner & R. L. Martens (Vol. Eds.), *Computer-supported collaborative learning: Vol. 3. What we know about CSCL _ And implementing it in higher education* (pp. 3–30). Boston, MA: Kluwer Academic Publishers.

Martens, R., Bastians, Th., & Gulikers, J. (2002). Leren met computergebaseerde authentieke taken: motivatie, gedrag en resultaten van studenten [Learning with computer-based authentic tasks: student motivation, behavior and results]. *Pedagogische Studiën*, 79(6), 469–482.

McDowell, L. (1995). The impact of innovative assessment on student learning. *Innovations in Education and Training International*, 32(4), 302–313.

Meyer, C. (1992). What's the difference between authentic and performance assesment? *Educational Leadership*, 49(8), 39–40.

Moerkerke, G., Doorten, M., & de Roode, F. A. (1999). *Constructie van toetsen voor competentiegericht curricula [Construction of assessments for competency-based curricula] (OTEC report 1999/W02)*. Heerlen, The Netherlands: Open Universiteit Nederland, Educational Technology Expertise Center.

Newmann, F. M. (1997). Authentic assessment in social studies: Standards and examples. In G. D. Phye (Ed.). *Handbook of classroom assessment: Learning, achievement, and adjustment* (pp. 359–380). San Diego, CA: Academic Press.

Newmann, F. M., & Wehlage, G. G. (1993). Five standards for authentic instruction. *Educational Leadership*, 50(7), 8–12.

Nicaise, M., Gibney, T., & Crane, M. (2000). Toward an understanding of authentic learning: Student perceptions of an authentic classroom. *Journal of Science Education and Technology*, 9, 79–94.

Petraglia, J. (1998). *Reality by design: The rhetoric and technology of authenticity in education*. Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Prodromou, L. (1995). The backwash effect: From testing to teaching. *ELT Journal*, 49(1), 13–25.

Reeves, T. C., & Okey, J. R. (1996). Alternative assessment for constructivist learning environments. In B.G. Wilson (Ed.). *Constructivist learning environments: Case studies in instructional design* (pp. 191–202). Englewood Cliffs, NJ: Educational Technology Publications.

Resnick, L. B. (1987). *Learning in school and out*. *Educational Leadership*, 16(9), 13–20.

Sambell, K., & McDowell, L. (1998). The construction of the hidden curriculum: Messages and meanings in the assessment of student learning. *Assessment and Evaluation in Higher Education*, 23(4), 391–402.

Sambell, K., McDowell, L., & Brown, S. (1997). But is it fair?: An exploratory study of student perceptions of the consequential validity of assessments. *Studies in Educational Evaluation*, 23(4), 349–371.

Savery, J., & Duffy, T. (1995). Problem based learning: An instructional model and its constructivist framework. *Educational Technology*, 35, 31–38.

Schnitzer, S. (1993). Designing and authentic assessment. *Educational Leadership*, 50(7), 32–35.

Segers, M., Dierick, S., & Dochy, F. (2001). Quality standards for new modes of assessment. An exploratory study of the consequential validity of the OverAll test. *European Journal of Psychology of Education*, 16(4), 569–586.

-
- Segers, M., Dochy, F., & Cascallar, E. (2003). *Optimising new modes of assessment: In search of qualities and standards*. Dordrecht, The Netherlands: Kluwer Academic Publishers.
- Segers, M., Dochy, F., & De Corte, E. (1999). Assessment practices and students' knowledge profiles in a problem-based curriculum. *Learning Environments Research*, 2, 191–213.
- Slavin, R. E. (1989). Research on cooperative learning: An international perspective. *Journal of Educational Research*, 33, 231–243.
- Sluijsmans, D. (2002). *Student involvement in assessment: the training of peer assessment skills*. Unpublished doctoral dissertation, Open University of the Netherlands, Heerlen, The Netherlands.
- Sweller, J., Van Merriënboer, J. J. G., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, 10(3), 251–296.
- Torrance, H. (1995). *Evaluating authentic assessment*. Buckingham, UK: Open University Press.
- Uhlenbeck, A. (2002). *The development of an assessment procedure for beginning teachers of English as a foreign language*. Unpublished doctoral dissertation, University of Leiden, Leiden, The Netherlands.
- Van Merriënboer, J. J. G. (1997). *Training complex cognitive skills: A four-component instructional design model for technical training*. Englewood Cliffs, NJ: Educational Technology Publications.
- Wiggins, G. (1989). Teaching to the (authentic) test. *Educational Leadership*, 46(7), 41–47.
- Wiggins, G. P. (1993). *Assessing student performance: Exploring the purpose and limits of testing*. San Francisco, CA: Jossey-Bass/Pfeiffer.